

Methodology for the United States Population Estimates by Age, Sex, Race, and Hispanic Origin (Vintage 2013): April 1, 2010 to July 1, 2013

Each year, the U.S. Census Bureau produces and publishes estimates of the population for each state and county, as well as the nation as a whole. We utilize administrative data from a number of sources to estimate 1) the change in population since the most recent decennial census and 2) the population for each year since the most recent decennial census. With each annual release of population estimates, the entire time series of estimates beginning on April 1, 2010 is revised and updated.

This document first describes the population estimates that we produce and publish. The second section describes the administrative data we utilize, our modifications and assumptions regarding those data, and the methods we use to estimate the components of population change. The final section describes the methods we use to produce the population estimates from the administrative data and the most recent decennial census data.

Population Estimates Data

For the nation, we release monthly estimates of the resident population by age, sex, race, and Hispanic origin. Additionally, we release national estimates by demographic characteristics of four other populations: the resident plus Armed Forces overseas, civilian, civilian noninstitutionalized, and household populations. Each of these four additional populations is based directly on the resident population.

For each state and county, we release annual estimates of the resident population by age, sex, race, and Hispanic origin. This document describes the production of the total resident population for states and counties.

Administrative Input Data

At the national level, the resident population is affected by births, deaths, and net international migration (NIM) only.

$$Pop\ Estimate = Base\ Pop + Births - Deaths + NIM$$

At the subnational level (e.g., states and counties), the resident population is affected by an additional component of population change: net internal, or domestic, migration (NDM).

$$Pop\ Estimate = Base\ Pop + Births - Deaths + NIM + NDM$$

Base Population

The enumerated resident population from the 2010 Census is the starting point for all post-2010 population estimates. We modify this enumerated population in two ways to produce the April 1, 2010 population estimates base.

First, we reconcile the 2010 Census race categories with the race categories that appear in our administrative data by recoding the “Some other race” responses in the 2010 Census to one or more of the five 1997 Office of Management and Budget (OMB) race categories: White; Black or African American; American Indian and Alaska Native; Asian; and Native Hawaiian and Other Pacific Islander.¹

Second, we update the population estimates base to reflect changes to the 2010 Census population due to the Count Question Resolution (CQR) program, legal boundary updates reported by January 1, 2013, and other geographic program revisions.²

Births

To estimate births, we utilize birth certificate data collected by the National Center for Health Statistics (NCHS). NCHS provided final individual birth records for births occurring before January 1, 2012 by date of birth, sex of child, residence and age of mother, and race and Hispanic origin of both mother and father. We modified these data in three main ways for use in estimates production.

First, not all states currently utilize the 1997 OMB race categories to request the parents’ race on birth certificates. Some states still record race according to the 1977 OMB race categories of White; Black; American Indian; Eskimo or Aleut; and Asian or Pacific Islander, under the “mark one race” scenario. Therefore, we must convert the parents’ reported race from the 1977 OMB race categories into the 1997 OMB race categories. We do this using the race bridging method designed by NCHS and the U.S. Census Bureau to make multiple-race and single-race data comparable.³

Second, birth certificates only require the race and Hispanic origin of the parents, not of the child directly. To impute the race and Hispanic origin of each child, we utilize the joint distribution of their parents’ race and origin and decennial census information on race and origin reporting within family households.

¹ The OMB standards are detailed in Office of Management and Budget, “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity” Notice, Vol. 62, No. 210, Thursday, October 30, 1997 <http://www.whitehouse.gov/omb/fedreg/1997standards.html>.

² For more information on the 2010 Census Count Question Resolution (CQR) program, see <http://2010.census.gov/2010census/about/cqr.php>.

³ For more information on the NCHS race-bridging factors, see http://www.cdc.gov/nchs/nvss/bridged_race.htm.

The third modification to the NCHS birth data accounts for inconsistencies between the race and Hispanic origin distributions of reported births and of the estimates base population under 1 year of age. We benchmark the distribution of births by race and Hispanic origin to be consistent with the 2010 Census population under 1 year of age. This allows us to have a “Census consistent” time series of births.

Once the birth certificate data from NCHS have been modified for estimates production, we calculate age, sex, race, and Hispanic origin-specific birth rates for 2012 using the Census Bureau’s population estimates for July 1, 2012. With these rates, we estimate the number of births by sex, race, and Hispanic origin occurring from April 1, 2010 to June 30, 2013. This process is described in the national estimates production section below.

For the production of state and county population estimates, we also estimate the annual number of births by sex, race, and Hispanic origin occurring in each county. We utilize the information on residence of mother from the calendar year 2011 NCHS data along with more current birth reports from the members of the Federal-State Cooperative for Population Estimates (FSCPE) to distribute the annual national-level births by sex, race, and Hispanic origin to each county.

Deaths

To estimate deaths, we utilize death data collected by NCHS. NCHS provided final individual death records for all deaths occurring before January 1, 2011 by residence, age, sex, race, and Hispanic origin of each decedent, as well as the place and date each death occurred. Since individual death records for 2011 were not available in time for production of the Vintage 2013 population estimates, deaths for 2011 were estimated using summary totals provided by NCHS for the nation, states, and counties. Deaths were distributed to age, sex, race, and Hispanic origin detail using the most current final mortality file available from NCHS (calendar year 2010).

As with birth data, the race of decedents from many states is still reported in the 1977 OMB race categories. Therefore, the first step in preparing death data for use in population estimates production is to convert the decedent’s reported race into the 1997 OMB race categories using the NCHS race bridging. The race-bridging process produces the full age, sex, race, and Hispanic origin distribution of monthly deaths for the nation.

Research shows that there are discrepancies in the reporting of age between census counts and death registration, most notably in the oldest ages. To address this issue, we redistribute all deaths occurring to the aggregate population 70 years and older by sex, race, and Hispanic origin to single years of age (70 to 99 and 100+ years) using life-table based death rates.⁴ We make no

⁴ To derive the death rates for the age 70 and older population, we employ life tables based on annual 2000-2010 NCHS mortality files and 2000-2010 intercensal population estimates prepared by the U.S. Census Bureau. The life tables are for males and females in five groups: Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, and non-Hispanic Asian and Pacific Islander.

additional adjustments, beyond race-bridging, to the deaths occurring to the population under 70 years of age.

Once the death certificate data from NCHS have been modified for estimates production, we calculated age, sex, race, and Hispanic origin-specific death rates for 2012 using the Census Bureau's population estimates for July 1, 2012. With these rates, we estimate the number of deaths by age, sex, race, and Hispanic origin occurring from April 1, 2010 to June 30, 2013. This process is described in the national estimates production section below.

For the production of state and county population estimates, we also estimate the annual number of deaths by age, sex, race, and Hispanic origin occurring in each county. We utilize the information on residence of decedent from the calendar year 2010 through 2011 NCHS data along with more current death reports from the members of the FSCPE to distribute the annual national-level deaths by age, sex, race, and Hispanic origin to each county.

Net International Migration

We estimate international migration in several parts: immigration of the foreign born, emigration of the foreign born, net migration between the United States and Puerto Rico, net migration of natives to and from the United States, and net movement of the Armed Forces population to and from the United States. For each component, we first estimate the total migration flow for the nation. To determine the state- and county-level age, sex, race, and Hispanic origin distribution of each component, proxy universes are developed that are assumed to be representative of the different components. The demographic characteristics and geographic distribution of these proxy universes are then applied to the totals for each component. For all components except movement of the Armed Forces population to and from the United States, state-level characteristics are based on the American Community Survey (ACS) three-year 2009-2011 file. County-level characteristics are based on data from the ACS five-year 2007-2011 file. County-level data are controlled to state-level data to ensure the component data sum as required. For the net movement of the Armed Forces population, demographic characteristics and state distributions are based on data collected by the Defense Manpower Data Center (DMDC), supplemented with data from the ACS five-year files. In Vintage 2013, supplemental data for 2010 come from the 2006-2010 ACS five-year file, data for 2011 from the 2007-2011 file, and data for 2012 and 2013 from the 2008-2012 file.

Immigration of the foreign born is estimated separately for Mexico and "All other countries" using the ACS question on residence one year ago. The foreign-born household population who indicated that they lived in Mexico in the prior year are considered immigrants from Mexico and the foreign-born household population who indicated that they lived abroad but not in Mexico are immigrants from "All other countries." The number of foreign-born migrants who entered the United States between April 2010 and June 2010, for both Mexico and "All other countries," is estimated as one quarter of the foreign-born household population in the 2010 ACS who reported living abroad one year ago. Because this question is asked only of those aged one and older, the estimate of foreign-born immigrants under the age of one is assumed to be equal to half the number of immigrants age one. Information from the 2011 ACS is used to estimate migration for

the July 2010 through June 2011 period. The estimate for the July 2011 through June 2012 period is estimated using data from the 2012 ACS. This estimate is held constant for the July 2012 through June 2013 period because more recent data are not available. The foreign-born household population whose place of birth was Mexico and whose year of entry was within five years of the survey year is used as the proxy universe to estimate the state- and county-level characteristics of foreign-born immigration from Mexico. The foreign-born household population whose place of birth was other than Mexico and whose year of entry was within five years of the survey year is used as the proxy universe to estimate the state- and county-level characteristics of foreign-born immigration from “All other countries.” Age in the ACS is modified for foreign-born immigrants to represent age at arrival to the United States.

Emigration of the foreign born from the United States is estimated separately for Mexico and “All other countries” using a residual method. For foreign-born emigration to Mexico, the foreign-born household population in Census 2000 whose place of birth was Mexico is aged forward (using NCHS life tables) to obtain the expected population in 2007, 2008, 2009, 2010, 2011, and 2012. The expected population is then compared to the population estimated in ACS 2007, ACS 2008, ACS 2009, ACS 2010, ACS 2011, and ACS 2012. Subtracting the estimated from the expected populations provides us with the residual, which serves as the basis for our emigration rates for the 2000 to 2007, 2000 to 2008, 2000 to 2009, 2000 to 2010, 2000 to 2011, and 2000 to 2012 time periods. This calculation is performed for two period-of-entry groups: the foreign born who entered the United States between 1990 and 1999, and the foreign born who entered before 1990. The method for estimating foreign-born emigration to “All other countries” is the same, except the foreign-born population whose place of birth was in a country other than Mexico is used in the residual calculations.

We then calculate three-year average rates for each period of entry group and apply the rates to the population at risk of emigrating each year to obtain estimates of emigration of the foreign-born population who entered the United States within the last ten years and of those who entered more than ten years ago. To produce estimates of foreign-born emigration from April 2010 through June 2010, the average of the rates from the 2000 to 2007, 2000 to 2008, and 2000 to 2009 residuals are applied by period of entry to ACS 2009. The estimates are divided by four to obtain estimates for the three-month period. For the estimates from July 2010 through June 2011, the average of the rates from the 2000 to 2008, 2000 to 2009, and 2000 to 2010 residuals are applied to ACS 2010. For the estimates from July 2011 through June 2012, the average of the rates from the 2000 to 2009, 2000 to 2010, and 2000 to 2011 residuals are applied to ACS 2011. For the estimates from July 2012 to June 2013, the average of the rates from the 2000 to 2010, 2000 to 2011, and 2010 to 2012 residuals are applied to ACS 2012. The proxy universe for foreign-born emigrants to Mexico who entered the United States within ten years of the estimate year is the foreign-born household population in the ACS whose place of birth was Mexico and who entered the United States within ten years of the survey year.

The proxy universe for foreign-born emigrants to Mexico who entered the United States more than ten years before the estimate year is the foreign-born household population in the ACS whose place of birth was Mexico and who entered the United States more than ten years before the survey year. The proxy universe for the foreign-born emigration to “All other countries” component is the same, except place of birth is restricted to countries other than Mexico.

Data from the ACS and the Puerto Rico Community Survey (PRCS) allow us to estimate the annual migration flows between the United States and Puerto Rico directly using the question on place of prior residence. People who indicated on the ACS that they lived in Puerto Rico one year ago are considered immigrants. People who indicated on the PRCS that they lived in the United States one year ago are considered emigrants. The proxy universe for the net migration between the United States and Puerto Rico is the population born in Puerto Rico whose year of entry was ten or fewer years before the survey year.

The net migration of the native born is based on research by Schachter (2008) using data from over 80 countries.⁵ This work compared estimates of the U.S. born or U.S. citizen population living overseas measured at two consecutive time periods and used the difference to develop estimates of net native migration. The proxy universe for the net native migration component is the native-born civilian population whose residence one year ago was either in a different state or abroad.

We derive the estimate of the net overseas movement of the Armed Forces population from data collected by DMDC. DMDC provides monthly tabulations of military personnel stationed or deployed outside the United States by age, sex, race, Hispanic origin, and individual branches of service within the Department of Defense. We assume that change in the overseas military population, excluding deaths, indicates movement of personnel in and out of the United States. To derive the estimates of net movement in certain counties, we apply the demographic and geographic distributions of the military population from ACS to DMDC total estimates by age, sex, race, and Hispanic origin.

Net Domestic Migration

We calculate domestic migration separately for three age groups (0-17, 18-64, and 65 and older) and two population universes (household and group quarters). To do this, we use information from tax exemptions, Medicare enrollment, and change in the group quarters population.

For the 0 to 64 year old household population (the 0-17 and 18-64 age groups), we use person-level data on filers and dependents aged 0 to 64 years from Federal income tax returns supplied by the Internal Revenue Service (IRS). We match two years of IRS tax return exemptions and compare the addresses to identify the number of individuals (represented by exemptions) who moved from one county to another between tax filings. Since every U.S. resident may not file or be claimed as an exemption on a tax return, we cannot use these data to directly estimate the number of county-to-county migrants. Instead, we calculate net domestic migration rates by subtracting the number of out-migrant exemptions from the in-migrant exemptions for each county to produce the number of IRS-based net migrant exemptions. Then, we divide the number of IRS-based net migrant exemptions by the sum of non-migrant exemptions and out-migrant exemptions for each county. We calculate these rates separately for two age groups within the

⁵ Schachter, Jason. 2008. "Estimating Native Emigration from the United States," Memorandum dated December 24, delivered to the U.S. Census Bureau.

population age 0 to 64 (0-17 and 18-64) so that we can create the state-level estimate of the population age 18 years and older for our annual release.

$$NDM Rate_{0-17} = \left(\frac{In Migrants_{0-17} - Out Migrants_{0-17}}{Non Migrants_{0-17} + Out Migrants_{0-17}} \right)$$

$$NDM Rate_{18-64} = \left(\frac{In Migrants_{18-64} - Out Migrants_{18-64}}{Non Migrants_{18-64} + Out Migrants_{18-64}} \right)$$

For the 65 years and older household population, we use annual Medicare enrollment data for each county from the Centers for Medicare and Medicaid (CMS). As with the IRS data, we know that not all U.S. residents aged 65 and older receive, or are eligible to receive, Medicare benefits. Therefore, we use the year-to-year change in the Medicare enrollment to calculate a domestic net migration rate. We assume that the year-to-year change in enrollment (benchmarked to the resident population on April 1, 2010 to adjust for undercoverage) represents the total change in the aged 65 and older population in each county. We utilize the estimates of deaths and international migration of the aged 65 and older population, as well as the number of individuals turning 65 years old in each period, to identify the amount of population change that, by default, must be due to net domestic migration. We calculate a Medicare-based net migration rate for each county by dividing the net domestic migration estimate by the total number of Medicare enrollees at the beginning of the time period.

$$NDM Rate_{65+} = \left(\frac{Change in Medicare Enrollment - (Pop turning 65 - Deaths_{65+} - NIM_{65+})}{Beginning Medicare Enrollment_{65+}} \right)$$

To estimate the change in the group quarters (GQ) population, we use data from two sources: 1) 2010 Census group quarters population by single year of age, sex, race, Hispanic origin, and facility type (both institutional and noninstitutional) for each subcounty area (e.g., cities, towns, etc.) and 2) a time series of total population in individual GQ facilities from the Group Quarters Report (GQR) prepared by the FSCPE members.⁶

From these data sources, we first estimate a time series of annual total GQ population by type by aggregating facility data to the subcounty level. We separately sum the GQ populations from the 2010 Census and the GQR by facility type for each estimate date in the time series. We then calculate a time series of GQ population by facility type by adding the year-to-year change indicated by the GQR data to the 2010 Census GQ populations by facility type.

Once the total GQ estimates are produced at the subcounty level, we sum the estimates to the county level. We use the 2010 Census GQ population by county and facility type as the proxy

⁶ The seven major GQ facility types utilized in estimate production are: correctional institutions, juvenile institutions, nursing homes, other institutional facilities, college dormitories, military housing, and other noninstitutional facilities.

universe to estimate the age, sex, race, and Hispanic origin distribution of the total county-level GQ population by facility type.

From these data, we aggregate the GQ population to the national level by age, sex, race, and Hispanic origin for use in the production of national estimates. We also aggregate the GQ population to the county level by age group (0 to 17 years, 18 to 64 years and 65 years and older) for use in the production of state and county total estimates.

Net domestic migration of the GQ population is represented by the annual change in the final GQ population estimates for each county and age group.

Production of Population Estimates

National Population by Age, Sex, Race, and Hispanic Origin

There are three main steps in the production of monthly national population estimates: estimating the quarterly national resident population; estimating the population for the other months of each quarter; and estimating the monthly population for the other four universes.

First, we calculate the births, deaths, and population for one quarter at a time using a cohort component method starting with the base population by age, sex, race, and Hispanic origin as of April 1, 2010. We begin by estimating the population at risk of giving birth or dying within the quarter of interest. Then, we multiply this population by the birth and death rates (by age, sex, race, and Hispanic origin) to estimate the births and deaths that occur in that quarter. Finally, we use those births and deaths, along with the net international migration component, to estimate the population in each birth cohort for the first day of the next quarter. We repeat this process one quarter at a time to the end of the time series.

With final quarterly estimates in hand, the second step in the process is to estimate the population for the other two months of each quarter. We assign the calculated quarterly births and deaths to specific months within each quarter based on the monthly distribution of births and deaths from 2011. Then, we use these monthly components of change and the balancing equation to estimate the population by age (not cohort), sex, race, and Hispanic origin for the remaining months in each year of the time series (August, September, November, December, February, March, May, and June).

The last step in the production of monthly national estimates is to calculate the four additional population universes by demographic characteristics. To calculate the resident plus Armed Forces overseas population, we add the monthly overseas military population, based on data collected by DMDC, to the resident population. To estimate the civilian population, we subtract the monthly resident military population, also estimated from data collected by DMDC, from the resident population. The civilian noninstitutionalized population is produced by subtracting the

institutionalized group quarters population from the civilian population.⁷ Finally, we estimate the household population by subtracting the total group quarters population from the resident population.

State and County Total Population

To produce annual state and county total population estimates, we begin by producing county-level estimates for three age groups: population aged 0 to 17, 18 to 64, and 65 and older.⁸ Starting with the county-specific base population by age group as of April 1, 2010, we calculate the population as of July 1, 2010 using a component of change method.

We first estimate the household population at risk of migrating into or out of each county between April 1 and June 30, 2010. Then, we multiply this population by the net domestic migration rates for the household population in each age group to estimate the county-specific net household domestic migration. We use the births, deaths, net international migration, and net domestic migration to estimate the resident population in each age group. Finally, we control the county population estimates by age group to the national resident population estimates for these age groups and repeat this process to estimate the annual July 1 resident population for each county.

The total resident population for each county is the sum of the populations in the three age groups. The resident population estimates and components of change for each state are simply the sum of the populations and components of change for its counties.

⁷ The institutionalized population is defined as people under formally authorized, supervised care or custody in institutions including correctional institutions, juvenile institutions, nursing homes, skilled nursing facilities, psychiatric hospitals, and facilities for the disabled.

⁸ We estimate the county population separately for ages 0 to 64 (including 0-17 and 18-64) and ages 65 and older because these groups are consistent with the populations at risk in our two sources of domestic migration data: IRS tax return data and Medicare enrollment data.